УЧЕНЫЕ ЗАПИСКИ КАЗАНСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
Том 151, кн. 2 Физико-математические науки *2009*

УДК 621.391.037.372

СТОЙКОСТЬ СТЕГАНОГРАФИЧЕСКИХ СИСТЕМ

Е.В. Разинков, Р.Х. Латыпов

Аннотация

Статья посвящена вопросам стойкости стеганографических систем. Дан обзор основных понятий стеганографии, рассмотрены различные подходы к стеганографической стойкости, факторы, влияющие на практическую стойкость стегосистем, а также основные виды стегоанализа, их плюсы и минусы.

Ключевые слова: стеганография, скрытая передача информации, стегоанализ, стеганографическая стойкость.

1. Введение

Стеганография – это наука и искусство скрытой передачи информации. Скрывается сам факт наличия обмена информацией. Это достигается путем встраивания сообщений в не вызывающий подозрений объект, называемый стеганографическим контейнером. В цифровой стеганографии в качестве контейнеров используются цифровые изображения, цифровое аудио и видео.

Основная задача стеганографии сформулирована в виде так называемой «проблемы заключенных». Двое заключенных, Алиса и Боб, сидят в разных камерах. Для того чтобы спланировать побег, им необходимо обмениваться информацией по открытому каналу связи, контролируемому нарушителем Евой. Для того, чтобы передать сообщение Бобу, Алиса встраивает его в стеганографический контейнер и передает результат встраивания (стего) по каналу связи [1].

В стеганографии рассматриваются три модели нарушителя [2].

1. Ева – пассивный нарушитель. В этом случае ее основная задача состоит в обнаружении самого факта наличия скрытой передачи информации, модифицировать стего у Евы возможности нет. Ева исследует перехваченное стего на предмет наличия скрытой информации. Если информация ею не обнаруживается, Ева пересылает стего Бобу. В противном случае канал связи блокируется.

2. Ева – активный нарушитель. В этом случае Ева модифицирует каждый перехваченный цифровой объект в попытках уничтожить возможное скрытое сообщение. В качестве атак могут применяться такие, например, преобразования стего, как сжатие с потерями или низкочастотная фильтрация.

3. Нарушитель называется злонамеренным (malicious), если его действия основываются на свойствах конкретного стеганографического алгоритма и направлены на тайное вмешательство в коммуникацию Алисы и Боба. Например, Ева может попытаться выдавать себя за Алису или Боба.

Как правило, в задачах скрытой передачи информации предполагается присутствие пассивного нарушителя.

Встраивание информации в стеганографический контейнер заключается в незначительной его модификации [3]. Внесенные изменения не должны обнаруживаться статистически и с помощью органов чувств человека. Если, к примеру, информация встраивается в цифровое изображение, то элементами контейнера, модифицируемыми при встраивании информации, могут быть интенсивность цветовых компонентов пикселов, коэффициенты различных преобразований.

1.1. Формальное определение стегосистемы. Дадим формальное определение стегосистемы. Пусть k – стеганографический ключ из множества K возможных стеганографических ключей, $k \in K$, M – множество возможных скрываемых сообщений, C – множество возможных стеганографических контейнеров. Стеганографическая система состоит из двух отображений: встраивающего отображения Emb и извлекающего отображения Ext:

$$\mathrm{Emb}: \ C \times K \times M \to C, \tag{1}$$

$$\mathrm{Ext}: \ C \times K \to M \tag{2}$$

таких, что $\mathrm{Ext}(\mathrm{Emb}(c,k,m),k) = m$ для всех $c \in C$, $k \in K$, $m \in M$. Встраивающее преобразование Emb генерирует на основе контейнера c, ключа k и сообщения m стегосообщение s, $s = \mathrm{Emb}(c,k,m)$.

2. Стеганографическая стойкость

2.1. Информационно-теоретическое определение стойкости. Пусть стеганографические контейнеры c имеют вероятностное распределение P_C, $P_C(c)$ – вероятность того, что Алисой будет выбран контейнер c для встраивания информации. Эта информация известна Еве.

Предполагается, что и стеганографический ключ k из множества K, и сообщение m из множества сообщений M выбираются равновероятно. На основе этой информации и распределения P_C с помощью формул полной вероятности можно получить P_S – вероятностное распределение получаемых в результате встраивания информации стего, $P_S(s)$ – вероятность того, что скрывающим преобразованием будет сгенерировано стего s.

Сравнение распределений P_C и P_S производится на основе расстояния Кульбака–Лейблера (относительной энтропии), которое определяется следующим образом[1]:

$$D\left(P_C||P_S\right) = \sum_{c \in C} P_C\left(c\right) \log \frac{P_C\left(c\right)}{P_S\left(c\right)}. \tag{3}$$

Относительная энтропия всегда неотрицательна и равна нулю тогда и только тогда, когда $P_C = P_S$. Относительная энтропия не является расстоянием в строгом математическом смысле (так как она асимметрична и не удовлетворяет неравенству треугольника).

Если $D(P_C||P_S) = 0$, то есть распределение создаваемых Алисой стего P_S совпадает с распределением стеганографических контейнеров P_C, известным нарушителю, то стегосистема является *абсолютно стойкой* [4], так как нарушитель не имеет возможности различить стего и контейнеры.

Если $D(P_C||P_S) \leq \varepsilon$, то стегосистема определяется как ε-стойкая [4]. Чем меньше значение ε, тем более стойкой к пассивным стегоаналитическим атакам является система.

2.2. Практическая стойкость. Оценить стойкость стегосистемы в теоретико-информационном смысле на практике не представляется возможным, поэтому вводится понятие стеганографической стойкости в практическом смысле.

Стеганографическая система называется *стойкой в практическом смысле*, если не существует стегоаналитического алгоритма, который был бы способен обнаруживать наличие скрытой информации. Таким образом, практическая стойкость

[1] Все логарифмы в данной работе берутся по основанию 2.

стегосистемы зависит от развития стегоаналитических методов и может понижаться с развитием методов стегоанализа.

Отметим факторы, влияющие на стойкость стеганографических систем:

- выбор стеганографического контейнера;
- способ изменения элементов контейнера;
- количество измененных элементов;
- правило выбора изменяемых элементов контейнера.

2.2.1. Выбор стеганографического контейнера. От выбора контейнера для встраивания информации сильно зависит вероятность обнаружения наличия скрытого сообщения нарушителем.

Рассмотрим случай использования цифровых изображений в качестве контейнеров. Не следует использовать для встраивания информации изображения с небольшим количеством цветов и изображения, созданные в графических редакторах. Следует также избегать использования изображения, которое ранее было в формате JPEG для встраивания информации в пространственной области. JPEG-компрессия оставляет «след» в изображении, который может быть обнаружен [5]. После же встраивания информации этот «след» сохранится, но существует способ определить то, что данное изображение не могло быть получено лишь в результате декомпрессии JPEG-изображения. Это, конечно же, вызовет подозрения.

2.2.2. Способ изменения элементов контейнера. При встраивании информации некоторые элементы контейнера (байты интенсивности цветовых компонент пикселов, JPEG-коэффициенты и др.) изменяются в соответствии с битами скрываемого сообщения.

Рассмотрим распространенный случай, когда в результате встраивания сообщение содержится в младших битах соответствующих элементов контейнера.

Есть два основных подхода к изменению младшего бита, если он не совпадает со встраиваемым:

1) изменить младший бит на противоположный (01000111 $\rightarrow$ 01000110, 00111010 $\rightarrow$ 00111011). Используется, например, в алгоритме Jsteg [6];

2) вычесть единицу из числа (01000111 $\rightarrow$ 01000110, 00111010 $\rightarrow$ 00111001). Используется, например, в алгоритме F5 [7].

Следует заметить, что использование первого способа во многих случаях существенно понижает стойкость стеганографической системы. В [6] описана гистограммная атака на алгоритм Jsteg, использующая именно это его свойство.

2.2.3. Количество измененных элементов. Очевидно, что чем меньше искажений внесено скрывающим преобразованием, тем ниже вероятность обнаружения скрытой информации.

Отношение количества переданных бит к количеству внесенных искажений называется эффективностью встраивания (embedding efficiency) [8]. Высокое значение этого параметра свидетельствует о возможности необнаруживаемой передачи сравнительно большого объема информации.

В качестве примера рассмотрим способ повышения эффективности встраивания, называемый матричным встраиванием (*matrix embedding*). Он был предложен в [7] как часть алгоритма F5, скрывающего информацию в изображениях формата JPEG.

Применение этого способа позволяет встроить k бит в n коэффициентов, где $n = 2^k - 1$, путем изменения только одного из этих коэффициентов.

Пусть H – матрица $k \times n$, столбцы которой представляют собой все ненулевые векторы длины k. Пусть x – вектор длины n из младших бит используемых коэффициентов, а m – вектор длины k, содержащий биты скрываемого сообщения.

Если $Hx = m$, то изменения вносить не требуется – сообщение уже содержится в младших битах коэффициентов. В противном случае вычисляется вектор $z = Hx - m$. Очевидно, что z совпадает с одним из столбцов матрицы H, например с j-м. Тогда сообщение m может быть встроено путем изменения j-го элемента вектора x. Получатель, извлекший последовательность x из принятого стего, вычисляет сообщение следующим образом: $m = Hy$.

Матричное встраивание позволяет встраивать k бит в $n = 2^k - 1$ коэффициентов с внесением в среднем $(1 - 1/2^k)$ изменений.

2.2.4. Правило выбора. Один из способов повышения устойчивости стеганографической системы – адаптивный выбор элементов стегоконтейнера для встраивания информации. К примеру, информация может встраиваться в наиболее зашумленные участки изображения, что усложняет ее обнаружение нарушителем.

Однако следует заметить, что адаптивное правило выбора элементов контейнера чаще всего не зависит или слабо зависит от секретного ключа [9]. Это свойство позволяет и нарушителю успешно применить это правило выбора, что может позволить ему провести успешную атаку на стегосистему.

Эта проблема может быть решена, если при встраивании сообщения используется информация, доступ к которой нарушитель получить не может. Такой подход к сокрытию, позволяя использовать преимущества адаптивного стеганографического преобразования, не понижает устойчивости системы, так как нарушитель не имеет возможности применить правило выбора элементов стегоконтейнера. Извлечение сообщения получателем возможно благодаря “wet paper codes”, предложенным в [10].

Например, в [11, 12] предложена стегосистема с адаптивным выбором элементов контейнера. Известно, что человеческий глаз не чувствителен к искажениям на границах объектов, поэтому для встраивания информации в этом методе используются пикселы, находящиеся на границах объектов на изображении.

3. Классификация стегоаналитических атак

Стегоанализ – это наука и искусство обнаружения скрытой информации или определения каких-либо параметров стегосистемы. В этом случае мы рассматриваем атаки Евы.

Стегоаналитические атаки различаются по используемым методам, по имеющейся у стегоаналитика информации, по получаемой в результате атаки информации. Есть два основных подхода к построению пассивных стегоаналитических атак: стегоанализ, основанный на контролируемом обучении, и статистический стегоанализ.

3.1. Стегоанализ, основанный на контролируемом обучении. Этот вид стегоанализа заключается в обучении классификатора на основе выборки, состоящей из большого количества стего и пустых стеганографических контейнеров. На вход классификатора подается вектор значений, вычисленных на основе стего и контейнеров из обучающей выборки [13, 14].

Плюсы этого вида стегоанализа:

- показывает хорошие результаты при правильном подборе параметров, подаваемых на вход классификатора;
- при обучении классификатора для какого-либо конкретного алгоритма может быть достигнуто достаточно точное обнаружение информации;
- нет необходимости в разработке статистических моделей – используется обучающая выборка;
- машинное обучение – хорошо исследованная область.

Минусы этого вида стегоанализа:

• для каждого конкретного стеганографического алгоритма требуется обучить отдельный классификатор, что непросто реализовать на практике;

• критически важен правильный выбор параметров, подаваемых на вход классификатора, в то время как четкая системная схема подбора этих параметров не разработана;

• некоторые параметры самого классификатора и параметры процесса обучения должны быть подобраны стегоаналитиком, а этот подбор часто может быть осуществлен только методом проб и ошибок;

• стегоаналитик не имеет возможности контролировать вероятности ошибок первого и второго рода;

• относящиеся к этому виду стегоанализа методы не способны, как правило, оценить стеганографический ключ или длину сообщения.

3.2. Статистический стегоанализ. Статистический стегоанализ направлен на обнаружение скрытого сообщения на основе исследования статистических закономерностей, нарушаемых скрывающим преобразованием. Атаки, относящиеся к этому виду стегоанализа, достаточно разнообразны [6, 13, 15] и используют различные свойства стеганографических контейнеров и алгоритмов.

Плюсы статистического стегоанализа:

• необходимый для статистического стегоанализа математический аппарат хорошо разработан и может быть напрямую применен при построении стегоаналитических атак;

• стегоаналитик имеет возможность контролировать вероятности ошибок первого или второго рода;

• есть возможность оценки стеганографического ключа, длины скрытого сообщения, местонахождения скрытой информации в стегоконтейнере.

Минусы статистического стегоанализа:

• эффективность статистического стегоанализа значительно уменьшается при наличии неточностей в используемых статистических моделях;

• статистическая нестационарность цифровых изображений вызывает значительные практические затруднения.

4. JPEG-стеганография и ее предел

Распространенность изображений в формате JPEG стала причиной особой актуальности алгоритмов JPEG-стеганографии и соответствующих стегоаналитических атак.

Все алгоритмы JPEG-стеганографии встраивают информацию путем модификации AC-коэффициентов JPEG-преобразования. Большинство существующих алгоритмов JPEG-стеганографии используют в качестве контейнеров изображения в формате JPEG, есть алгоритмы, которые используют несжатые изображения, встраивая информацию в процессе проведения JPEG-преобразования и учитывая отбрасываемую информацию [9].

Использование нулевых JPEG-коэффициентов приводит к значительному понижению стойкости стегосистемы, поэтому для встраивания информации используются только ненулевые AC-коэффициенты.

Исследования показывают, что на данный момент лучшие алгоритмы JPEG-стеганографии могут обеспечить стойкую к пассивным атакам передачу информации при пропускной способности, не превышающей 0.05 бит на ненулевой AC-

коэффициент (стойкой в данном случае считается стегосистема, для которой полусумма вероятностей ошибок первого и второго рода превышает 0.4) [16].

Заключение

Стеганографическая стойкость – ключевое понятие стеганографии, определяющее место стеганографических методов в решении задач защиты информации [17–24]. В настоящей статье рассмотрены базовые понятия стеганографии, понятия информационно-теоретической и практической стойкости стегосистем, способы повышения практической стойкости, классификация стегоаналитических атак.

Summary

E.V. Razinkov, R.Kh. Latypov. Security of Steganographic Systems.

This paper regards the security of steganographic systems. Two main approaches to steganographic security are discussed. Two types of steganalysis are examined, advantages and disadvantages of each type of steganalysis are considered.

Key words: steganography, steganalysis, information hiding, steganographic security.

Литература

1. *Simmons G.J.* The Prisoners' Problem and the Subliminal Channel // Proc. of Crypto'83 / Ed. D. Chaum. – N. Y.: Plenum Press, 1984. – P. 51-67.
2. *Cox I.J., Miller M.L., Bloom J.A., Fridrich J., K. Ton* Digital Watermarking and Steganography. – Elsevier, 2008. – 593 p.
3. *Wayner P.* Disappearing Cryptography. Information Hiding: Steganography and Watermarking. – Elsevier, 2002. – 413 p.
4. *Cachin C.* An Information-Theoretic Model for Steganography // Proc. of 2nd Workshop on Information Hiding / Ed. D. Aucsmith. Lecture Notes in Computer Science. – N. Y.: Springer-Verlag, 1998. – V. 1525. – P. 306–318.
5. *Fridrich J., Goljan M., Du R.* Steganalysis Based on JPEG Compatibility // Special session on Theoretical and Practical Issues in Digital Watermarking and Data Hiding, SPIE Multimedia Systems and Applications IV. – Denver, CO, 2001. – P. 275-280.
6. *Lee K., Westfeld A., Lee S.*Generalised Category Attack – Improving Histogram-Based Attack on JPEG LSB Embedding // Information Hiding. 9th Int. Workshop. Lecture Notes in Computer Science. – Springer-Verlag, 2007. – V. 4567. P. 378–391.
7. *Westfield A.* F5 – A Steganographic Algorithm: High Capacity Despite Better Steganalysis // Information Hiding. 4th Intern. Workshop. Lecture Notes in Computer Science. – Berlin, Heidelberg, New York: Springer-Verlag, 2001. – V. 2137. – C. 289–302.
8. *Fridrich J., Lisonek P., Soukal D.* On Steganographic Embedding Efficiency // Information Hiding. 8th Int. Workshop. Lecture Notes in Computer Science. – Springer-Verlag, 2007. – V. 4437. – P. 282–296.
9. *Fridrich J., Goljan M., Soukal D.* Perturbed Quantization Steganography // ACM Multimedia & Security Journal. – 2005. – V. 11, No 2. – P. 98–107.
10. *Fridrich J., Goljan M., Lisonek P., Soukal D.* Writing on Wet Paper // IEEE Trans. On Sig. Proc. Special Issue on Media Security / Eds. T. Kalker, P. Moulin. – 2005. – V. 53. – P. 3923–3935.
11. *Разников Е.В., Латыпов Р.Х.* Скрытая передача информации с использованием границ объектов // Учен. зап. Казан. ун-та. Сер. Физ.-матем. науки. – 2007. – Т. 149, кн. 2. – С. 128–137.

12. *Razinkov E.V., Latypov R.Kh.* Image Steganography Technique Using Objects Outlines // Proc. of the IEEE Systems, Man and Cybernetic Society 6th Conference on Cybernetic Systems. – Dublin, 2007. – P. 46–50.

13. *Fridrich J., Goljan M.* Practical Steganalysis of Digital Images – State of the Art // Proc. SPIE Electronic Imaging, Security and Watermarking of Multimedia Contents. – San Jose, CA, USA, 2002. – P. 1–13.

14. *Farid H.* Detecting Steganographic Message in Digital Images: Report TR2001-412. – Hanover, NH: Dartmouth College, 2001.

15. *Chandramouli R., Subbalakshmi K.* Current Trends in Steganalysis: A Critical Survey // IEEE Int. Conf. on Control, Automation, Robotics and Vision, ICARCV. – 2004. – V. 2. – P. 964–967.

16. *Fridrich J., Pevny T., Kodovsky J.* Statistically Undetectable JPEG Steganography: Dead Ends, Challenges, and Opportunities // Proc. ACM MM&S Workshop / Eds. J. Dittmann, J. Fridrich. – Dallas, TX, 2007. – P. 3–14.

17. *Sullivan K., Madhow U., Chandrasekaran S., Manjunath B.* Steganalysis for Markov Cover Data With Applications to Images // IEEE Transactions on Information Forensics and Security. – 2006. V. 1, No 2. – P. 275–287.

18. *Sencar H.T., Ramkumar M., Akansu A.N.* Data Hiding Fundamentals and Applications: Content Security in Digital Multimedia. – Orlando, FL, USA: Acad. Press, 2004. – 252 p.

19. *Duric Z., Jacobs M., Jajodia S.* Information Hiding: Steganography and Steganalysis // Handbook of Statistics: Data Mining and Data Visualization. – 2005. – V. 24. – P. 171–188.

20. *Fridrich J., Goljan M., Hogea D.* Steganalysis of JPEG Images: Breaking the F5 Algorithm // Revised Papers from the 5th Int. Workshop on Information Hiding. – 2002. – P. 310–323.

21. *Chandramouli R., Kharrazi M., Memon N.D.* Image steganography and steganalysis: Concepts and practice // Digital Watermarking, 2nd Int. Workshop, IWDW 2003, Seoul, Korea, 2003. Lecture Notes in Computer Science. – N. Y.: Springer-Verlag, 2004. – V. 2939. – P. 35–49.

22. *R. Chandramouli* A mathematical framework for active steganalysis // ACM Multimedia Systems. – 2003. – V. 9, No 3. – P. 303–311.

23. *Fridrich J., Goljan M., Hogea D.* Attacking the OutGuess // Proc. ACM Workshop Multimedia and Security 2002. – N. Y.: ACM Press, 2002. – URL: http://faculty.ksu.edu.sa/ghazy/Steg/References/Ref11.pdf.

24. *Fridrich J., Goljan M., Soukal D.* Searching for the Stego Key // Proc. SPIE-Security, Steganography and Watermarking of Multimedia Contents VI, Electronic Imaging. – San Jose, CA, 2004. – V. 5306. – P. 70–82.

Поступила в редакцию
23.03.09

Разинков Евгений Викторович – аспирант факультета ВМК Казанского государственного университета.

E-mail: *razinkov@steganography.ru*

Латыпов Рустам Хафизович – доктор технических наук, профессор, декан факультета ВМК Казанского государственного университета.

E-mail: *Roustam.Latypov@ksu.ru*